The combined impact of metabolic gene polymorphisms on elite endurance athlete status and related phenotypes

Ildus I. Ahmetov[1,2,*], Alun G. Williams[3], Daniil V. Popov[1], Ekaterina V. Lyubaeva[1], Albina M. Hakimullina[2], Olga N. Fedotovskaya[2], Irina A. Mozhayskaya[2], Olga L. Vinogradova[1], Irina V. Astratenkova[2]; Hugh E. Montgomery[4] and Viktor A. Rogozkin[2]

[1]Laboratory of Muscle Performance, SSC RF Institute for Biomedical Problems of the Russian Academy of Sciences, 123007, Moscow, 76A Khoroshevskoe chaussee, Russia. [2]Sports Genetics Laboratory, St Petersburg Research Institute of Physical Culture, 191040, St Petersburg, 56 Ligovsky Avenue, Russia. [3]Department of Exercise and Sport Science, Manchester Metropolitan University, Hassall Road, Alsager, Cheshire ST7 2HL, UK. [4]UCL Institute for Human Health and Performance, London, N19 5LW, UK.

*To whom correspondence should be addressed. Tel: +79655867625; Fax: +74991952253; Email: genoterra@mail.ru

ABSTRACT

Endurance performance is a complex phenotype, subject to the influence of both environmental and genetic factors. Whilst the last decade has seen a variety of specific genetic factors proposed, many in metabolic pathways, each is likely to make a limited contribution to an 'elite' phenotype: it seems more likely that such status depends upon the simultaneous presence of multiple such variants. The aim of the study was to investigate individually and in combination the association of common metabolic gene polymorphisms with endurance athlete status, the proportion of slow-twitch muscle fibers and maximal oxygen consumption. A total of 1,423 Russian athletes and 1,132 controls were genotyped for 15 gene polymorphisms, most of which were previously reported to be associated with athlete status or related intermediate phenotypes. Muscle fiber composition of *m. vastus lateralis* in 45 healthy men was determined by immunohistochemistry. Maximal oxygen consumption of 50 male rowers of national competitive standard was determined during an incremental test to exhaustion on a rowing ergometer. Ten 'endurance alleles' (*NFATC4* Gly160, *PPARA* rs4253778 G, *PPARD* rs2016520 C, *PPARGC1A* Gly482, *PPARGC1B* 203Pro, *PPP3R1* promoter 5I, *TFAM* 12Thr, *UCP2* 55Val, *UCP3* rs1800849 T and *VEGFA* rs2010963 C) were first identified showing discrete associations with elite endurance athlete status. Next, to assess the combined impact of all 10 gene polymorphisms, all athletes were classified according to the number of 'endurance' alleles they possessed. The proportion of subjects with a high (≥ 9) number of 'endurance' alleles was greater in the best endurance athletes compared to controls (85.7% vs 37.8%, $P = 7.6 \times 10^{-6}$). The number of 'endurance' alleles was shown to be positively correlated ($r = 0.50$; $P = 4.0 \times 10^{-4}$) with the proportion of fatigue-resistant slow-twitch fibers, and with maximal oxygen consumption ($r = 0.46$; $P = 7.0 \times 10^{-4}$). These data suggest that the likelihood of becoming an elite endurance athlete depends on the carriage of a high number of endurance-related alleles.

INTRODUCTION

The capacity to perform endurance exercise is influenced by a number of factors, many relating to cellular metabolism and cardiovascular function (Moore 1998; Bassett and Howley 2000; Flück 2006). These include the proportion of slow-twitch fibers in skeletal muscle (Abernethy et al. 2000) and factors such as maximal cardiac output which underlie the maximal rate of oxygen consumption (VO_{2max}) (Bassett and Howley 2000). Such intermediate phenotypes are under strong genetic influence. Indeed, some 40-50% of the variance in the proportion of slow-twitch fibers in human muscles seems genetically determined (Simoneau and Bouchard 1995), whilst VO_{2max} also has high heritability ($h^2 \sim 0.5$-0.6) (Bouchard et al. 1998, 1999). Athlete status itself is also a heritable trait: one report has suggested that around two-thirds of the variance in athlete status is explained by additive genetic factors, with the remaining variance attributed to non-shared environmental factors (De Moor et al. 2007).

In the decade since the *ACE* (angiotensin I converting enzyme) gene was first proposed to be a 'human gene for physical performance' (Gayagay et al. 1998; Montgomery et al. 1998), there have been numerous studies examining the effects of *ACE* and other genes on athletic status (Ahmetov and Rogozkin 2009; Bray et al. 2009). In several cases, these gene variants have been associated with aerobic capacity and muscle fiber composition as endurance-related traits. We have previously reported seven gene variants (*PPARA* rs4253778, *PPARD* rs2016520, *PPARGC1A* rs8192678, and *PPP3R1* promoter 5I/5D, *UCP2* rs660339, *UCP3* rs1800849 and *VEGFA* rs2010963) of inter-related metabolic pathways to be associated with elite endurance athlete status in Russians using a case-control study design (Ahmetov et al. 2006, 2007a, b, 2008a, b, c). These findings were in agreement with the data from several reports where relevant intermediate phenotypes were studied (Buemann et al. 2001; Halsall et al. 2001; Jamshidi et al. 2002; Poirier et al. 2003; Lucia et al. 2005; Tang et al. 2005; Vanttinen et al. 2005; Prior et al. 2006; Ahmetov et al. 2006, 2007b, 2008a, b, c). Of the seven gene variants we have examined previously, the *PPARA*, *PPARD*, *PPARGC1A* and *PPP3R1* genes code for transcription factors and coactivators, while *UCP2*, *UCP3* and *VEGFA* represent their target genes (Table 1).

Despite the clear role of genetics in human athletic performance, there is little unequivocal evidence in support of a specific genetic variant with a *major* effect on a relevant performance phenotype, at least across the normal range of human trait distributions. This may be because complex traits are fundamentally polygenic (many genes with small effects), or because researchers have failed to take into consideration the full range of environmental effects, or both (Brutsaert and Parra 2006). To date, the concept that endurance performance is likely to be determined by the simultaneous presence of many advantageous genetic variants has only been addressed in principle (Williams and

Folland 2008): few studies have yet sought to define or quantify the impact of multiple (i.e. more than two) genotype combinations that influence human physical performance and none have attempted this for more than seven genetic variants (Williams et al. 2004; Saunders et al. 2006; Gonzalez-Freire et al. 2008; Muniesa et al. 2008; Gómez-Gallego et al. 2009; Ruiz et al. 2009). We have thus addressed this issue, in a study focused on gene variants in metabolic pathways (i.e. genes primarily involved in ATP, glucose, insulin and lipid metabolism, mitochondrial biogenesis, thermogenesis, regulation of muscle fiber type composition and angiogenesis). The aim of the present study was therefore to investigate individually, and in combination, the associations of multiple common metabolic gene polymorphisms with endurance athlete status, the proportion of slow-twitch muscle fibers and maximal oxygen consumption.

MATERIALS AND METHODS

Study participants. In total, 1,423 male (n = 998) and female (n = 425) Russian athletes (age 24.4 ± 0.3 yr) of regional or national competitive standard were recruited from the following sports: alpine skiing (n = 13), artistic gymnastics (n = 55), basketball (n = 33), biathlon (n = 29), bodybuilding (n = 74), boxing (n = 30), tennis (n = 29), cross-country skiing (n = 142), soccer (n = 42), ice hockey (n = 17), jumping events (athletics) (n = 12), kayaking (n = 33), powerlifting (n = 26), race walking (n = 23), road cycling (n = 108), rowing (n = 191), running (n = 134), skating (n = 109), ski jumping (n = 14), swimming (n = 106), throwing events (n = 17), triathlon (n = 29), weightlifting (n = 61) and wrestling (n = 96). The athletes were prospectively stratified into five groups according to event duration and distance, covering a spectrum from the more endurance-oriented to the more power-oriented. The first three groups included very long (n = 288; race duration > 30 min; 5-25 km swimmers (n = 21), 15-50 km cross-country skiers (n = 78), biathletes, race walkers, road cyclists, triathletes), long (n = 290; race duration 5-30 min; 3-10 km runners (n = 5), 5-10 km skaters (n = 4), 5-10 km cross-country skiers (n = 64), 800-1500 m swimmers (n = 26), rowers) and middle (n = 116; race duration 45 s – 5 min; 200-400 m swimmers (n = 24), 800-1500 m runners (n = 7), 1500-3000 m skaters (n = 52), 500-1000 m kayakers) distance athletes (‘long endurance group’, ‘middle endurance group’ and ‘short endurance group’, respectively). The athletes of the long and middle endurance groups were also considered as ‘predominantly endurance-oriented’ athletes. The fourth group (n = 248; basketball players, boxers, ice hockey players, soccer players, tennis players, wrestlers) comprised athletes whose sports utilized mixed anaerobic and aerobic energy production (‘mixed group’). The fifth group (n = 481; 50-100 m swimmers (n = 35), 100-400 m runners (n = 122), 500-1000 m skaters (n = 52), alpine skiers, artistic gymnasts, bodybuilders, jumpers, powerlifters, ski jumpers, throwers and weightlifters) included sprint and strength athletes with predominantly

anaerobic energy production ('power group'). Age, height and body mass of athletes from different groups are presented in Table 2. There were 235 athletes classified as 'elite' (ranked in the top 10 nationally), of which 58 athletes were 'top elite' athletes (prize winners of the World and European Championships, World Cups and Olympic Games). There were 404 athletes classified as 'sub-elite' (participants in international competitions). The others (n = 784) were classified as 'non-elite' athletes, being regional competitors with no less than four years experience participating in their sports.

Controls were 1,132 healthy unrelated citizens of St Petersburg, Moscow, Naberezhniye Chelny and Surgut (537 males and 595 females; 17.2 ± 0.2 yr) without any competitive sport experience. Geographic ancestry of the athletes and control groups was self-reported. The athletes and control groups were all Caucasians: 89.3% of athletes and 84.5% of controls were Russians, while other frequently reported nationalities were Tatars and Ukrainians. The University of St Petersburg Ethics Committee approved the study and written informed consent was obtained from each participant.

Additionally, 45 physically active healthy men (23.5 ± 0.4 yr; height 180.0 ± 2.4 cm, mass 72.7 ± 1.4 kg) and 50 male rowers (20.6 ± 0.3 yr; 191.8 ± 0.8 cm, 87.4 ± 1.3 kg; VO_{2max} 55.7 ± 0.9 ml/min/kg) of the national competitive standard gave their informed consent to participate in studies of muscle fiber proportion and aerobic power, respectively. Both studies were approved by the Physiological Division of the Russian National Bioethics Committee.

Genotyping. Molecular genetic analysis was performed with DNA samples obtained from epithelial mouth cells by alkaline extraction or using a DNK-sorb-A sorbent kit according to the manufacturer's instruction (Central Research Institute of Epidemiology, Moscow, Russia), depending on the method of sample collection (buccal swab or scrape). Genotyping for 15 gene polymorphisms was performed by PCR on multicanal amplificator Tercyk (DNA Technology, Moscow, Russia) and restriction enzyme digestion (Supplementary Methods). All genotyping analyses were conducted blind to subject identity. All polymorphisms were genotyped in duplicate, with a further 10% of samples for all polymorphisms genotyped in a separate run as a quality assurance measure, with 100% agreement between runs.

For our genotype combination analysis, we used the genotypic data of our recent single-gene case-control and cross-sectional studies (Ahmetov et al. 2006, 2007a, b, 2008a, b, c) but also supplemented those data sets with new data due to the availability of increased sample sizes (by ~ 50%). In total, we genotyped 1,423 Russian athletes and 1,132 controls for the seven gene variants investigated previously (*PPARA* rs4253778, *PPARD* rs2016520, *PPARGC1A* rs8192678, and *PPP3R1*

promoter 5I/5D, *UCP2* rs660339, *UCP3* rs1800849 and *VEGFA* rs2010963). We also genotyped the athletes and controls for eight additional candidate gene polymorphisms considered likely to influence endurance performance, namely *ACE* Alu I/D, *AMPD1* rs17602729, *HIF1A* rs11549465, *NFATC4* rs2229309, *PPARG* rs1801282, *PPARGC1B* rs7732671, *TFAM* rs1937 and *VEGFA* rs699947 (Table 1).

Immunohistochemistry. Samples of *m. vastus lateralis* of 45 physically active healthy men were obtained with the Bergstrom needle biopsy procedure under local anesthesia with 1% lidocaine solution. Prior to analysis, samples were frozen in liquid nitrogen and stored at < -80°C. Serial sections (10 μm) were prepared using a cryostat and microtome at -20°C, with sections then mounted on slides. The immunoperoxidase technique was employed for immunohistochemical identification of myosin isoforms. Antibodies against the slow (MHCs) and fast (MHCf) myosin isoforms were used (clones NCL-MHCf (a+b) and NCL-MHCs (Novocastra Laboratories, Newcastle, UK)). Sections incubated without primary antibodies were to detect nonspecific staining. The antigen-antibody marking was intensified with the Vectastain ABC kit (Vector Labs Inc., Burlingame, CA, USA) to visualize the diaminebenzidine peroxidase reaction. Fiber distribution was expressed as a ratio of the number of fibers of each type in a section to the total number of fibers. All fibers (200-300) were measured in each section. The cross-sectional area (CSA) was determined for at least 100 fibers of each type using an image analysis system QUANTIMET-500 (Leica, Cambridge Ltd, Cambridge, UK) and a color digital video camera JVC TK-1280E (Tokyo, Japan; image resolution 720 x 512 pixels with 8 bit/pixel). Sections used for analysis were prepared and stained all together with Sigma (St. Louis, MO, USA) reagents. The proportion of the area (%) occupied by slow-twitch and fast-twitch fibers for each subject was also determined.

Aerobic power measurement. Aerobic power was determined using an incremental test to exhaustion on a rowing ergometer PM 3 (Concept II, Morrisville, Vermont, USA). The initial workload was 150 W. The duration of exercise at each workload was 3 minutes, with a 30 s rest period between increments of 50 W. Oxygen consumption (VO_2) was determined breath by breath using a MetaMax 3B gas analysis system (Cortex, Leipzig, Germany). Using the MetaMax system, O_2 and CO_2 contents were measured using an electro-chemical cell and nondispersive infrared sensor, respectively, and air flow was measured using a turbine transducer (Triple V). Two-point gas calibrations (first gas – 15% O_2, 5% CO_2; second gas – ambient air) were performed daily. A one-point gas calibration with ambient air was performed before each test as well as a flow transducer calibration using a 3 L syringe

(Hans Rudolph, Kansas City, USA). The criteria used to confirm a maximal test were a decrease in power of more than 30 W from the target power despite strong verbal encouragement and a respiratory exchange ratio greater than 1.1 before cessation of exercise. Maximal oxygen consumption (VO_{2max}) was recorded as the highest mean value observed over a 30 s period.

Statistical analysis. Genotype distribution and allele frequencies between each of the five groups of athletes (long-endurance, middle-endurance, short-endurance, mixed and power) and controls were compared using χ^2 tests. Within the four groups of athletes differing from controls in genotype distribution and allele frequencies (long-endurance, middle-endurance, short-endurance and mixed), the percentages of participants with high numbers of 'endurance' alleles were analyzed for linear trend across achievement level (top elite, elite, sub-elite, non-elite). Because the 'endurance' allele data were not distributed parametrically, Spearman's (non-parametric) correlations were used to assess the relationships between the physiological phenotypes (muscle fiber characteristics and maximal oxygen consumption) and the numbers of 'endurance' alleles possessed. Differences in aerobic power and muscle fiber composition between groups possessing high and low numbers of 'endurance' alleles were analyzed using unpaired *t* tests. All values are means ± SE (s.e.m.). *P* values < 0.05 were considered statistically significant. Bonferroni's correction for multiple testing was performed by multiplying the *P* value with the number of tests where appropriate. Statistical analyses were conducted using GrathPad Instat and StatistiXL v. 1.8 software.

RESULTS

Genotype distributions of 15 gene polymorphisms in the control group and amongst all athletes were in Hardy-Weinberg equilibrium. There were no significant differences in genotype and allele frequencies between males and females amongst athletes and controls, nor between the different Caucasian groups or citizens (controls) of different towns (data not shown). Therefore, for the main analyses we used the combined data (i.e. combined groups of male and female Caucasians, independent of precise nationality).

We first compared genotype distribution and allele frequencies of 15 gene variants between athletes (n = 287) of long endurance and middle endurance groups of national competitive standard (elite and sub-elite) and controls, and confirmed that seven 'endurance' alleles (i.e. the alleles individually associated with endurance athlete status or related phenotypes from previous published literature: *PPARA* G, *PPARD* C, *PPARGC1A* Gly482, *PPP3R1* 5I, *UCP2* 55Val, *UCP3* T, *VEGFA* rs2010963 C) were significantly overrepresented in our cohort of endurance-oriented athletes (Table 3). The frequencies of a further three alleles (*NFATC4* Gly160, *PPARGC1B* 203Pro, *TFAM* 12Thr)

were also higher in endurance-oriented athletes compared to controls, and as this was found for the first time, they can be considered as novel 'endurance' alleles. Thus, we had 10 statistically significant associations between endurance athlete status and allele frequency, initially. The associations of six of those alleles (*NFATC4* Gly160, *PPARGC1A* Gly482, *TFAM* 12Thr, *UCP2* 55Val, *UCP3* rs1800849 T and *VEGFA* rs2010963 C) with endurance athlete status remained statistically significant after Bonferroni correction for multiple testing, whilst the associations between endurance athlete status and the other four alleles (*PPARA* rs4253778 G, *PPARD* rs2016520 C, *PPARGC1B* 203Pro and *PPP3R1* promoter 5I) became non-significant. However, given that the latter four polymorphisms have recently been reported to be associated with endurance athlete status and/or related phenotypes (Jamshidi et al. 2002; Ahmetov et al. 2006, 2007a,b; Ling et al. 2007; Ahmetov et al. 2008b; Eynon et al. 2009), we felt justified in retaining all 10 alleles for further analysis. The other five polymorphisms (showing no association with endurance athlete status in our cohort; see Table 3) were not included in any further analyses: *ACE* Alu I/D, *AMPD1* C34T (rs17602729), *HIF1A* Pro582Ser (rs11549465), *PPARG* Pro12Ala (rs1801282) and *VEGFA* C-2578A (rs699947).

Next, to assess the combined impact of all 10 gene polymorphisms that were retained for further analysis, we classified all athletes and controls according to the number of 'endurance' polymorphic alleles (see Supplementary Table 1) they possessed (e.g. carriers of *NFATC4* Ala/Ala, *PPARA* CC, *PPARD* TT, *PPARGC1A* Ser/Ser, *PPARGC1B* Ala/Ala, *PPP3R1* 5D/5D, *TFAM* Ser/Ser, *UCP2* Ala/Ala, *UCP3* CC, *VEGFA* GG genotype had 0 'endurance' alleles, and subjects with *NFATC4* Gly/Gly, *PPARA* GG, *PPARD* CC, *PPARGC1A* Gly/Gly, *PPARGC1B* Pro/Pro, *PPP3R1* 5I/5I, *TFAM* Thr/Thr, *UCP2* Val/Val, *UCP3* TT, *VEGFA* CC genotype had 20 'endurance' alleles). However, neither athletes nor controls in our study had the minimal (i.e. 0) or maximal (i.e. 20) 'endurance' score (because of low minor allele frequencies for some gene polymorphisms). Accordingly, the 'endurance' score ranged from 3 to 13 for controls, and from 5 to 14 for the predominantly endurance-oriented athletes (athletes of long endurance and middle endurance groups; n = 578). The most frequently observed number of 'endurance' alleles in controls and endurance-oriented athletes was 8 (21.7%) and 9 (24.6%) respectively. On this basis, we classified all subjects into two groups as having a low (≤ 8) or high (≥ 9) number of 'endurance' alleles.

Figure 1 shows the distribution of athletes with high numbers of 'endurance' alleles stratified by endurance/power orientation and competitive standard in comparison with controls. When compared to controls, a greater proportion of international level athletes in the long-endurance, middle-endurance, short-endurance and mixed groups exhibited a high (≥ 9) number of 'endurance' alleles. Furthermore, the largest difference was seen when the top elite predominantly endurance-oriented athletes only (n = 21, not shown separately in Fig. 1) were compared to controls (85.7% vs

37.8%, P = 7.6 x 10^{-6}). As expected, the proportion of 'power' athletes with a high number of 'endurance' alleles did not differ significantly from controls.

The proportion of athletes with a high number of 'endurance' alleles was also significantly associated with elite athlete status within event groups. Accordingly, the proportions of elite athletes with high numbers of 'endurance' alleles in the long-endurance (76.4% vs. 56.6%; P = 0.010) and middle-endurance (71.7% vs. 44.1%; P = 0.003) groups were significantly higher in comparison with non-elite athletes. Furthermore, linear trends (assessed using χ^2 tests and the achievement level (non-elite → sub-elite → elite) as the categorical variable) were evident within the long-endurance (56.6% → 75.0% → 76.4%; P = 0.002) and middle-endurance (44.1% → 62.4% → 71.7%; P = 0.0004) groups, with tendencies for linear trends within the short-endurance (46.2% → 60.0% → 70.5%; P = 0.067) and mixed groups (45.6% → 62.9% → 60.0%; P = 0.058) (Fig. 1).

We also examined the combined impact of the 10 gene polymorphisms on the two intermediate endurance phenotypes we assessed, namely the proportion of slow-twitch muscle fibers in *m. vastus lateralis* of physically active healthy men (n = 45) and maximal oxygen consumption in rowers of the national competitive standard (VO_{2max} 55.7 ± 0.9 ml/min/kg; n = 50). The number of 'endurance' alleles positively correlated with the proportion of slow-twitch fibers (r = 0.50; P = 4.0 x 10^{-4} for Spearman correlation). For men with high numbers of 'endurance' alleles (n = 26) compared to those with low numbers of 'endurance' alleles (n = 19), there was a greater number of slow-twitch fibers in the *m. vastus lateralis* (56.1 ± 1.8% vs. 43.8 ± 2.2%; P = 1.0 x 10^{-4}) and a higher proportion of area occupied by those fibers (50.0% vs. 41.8%; P = 0.033).

The maximal oxygen consumption of rowers was also genotype-dependent. We observed a moderate correlation (r = 0.46; P = 7.0 x 10^{-4}) between VO_{2max} values of athletes and their carriage of 'endurance' alleles number. Those rowers having high numbers of 'endurance' alleles (n = 25; VO_{2max} 58.6 ± 1.2 ml/min/kg) showed significantly higher maximal oxygen consumption than their counterparts with low numbers of 'endurance' alleles (n = 25; VO_{2max} 52.8 ± 1.0 ml/min/kg; P = 6.0 x 10^{-4}).

DISCUSSION

Ten common polymorphisms have been shown here to contribute to a complex genetic profile of elite endurance athlete status – both in a comparison between elite athletes and non-athlete controls, and within the athlete cohort by level of competitive achievement. Seven of the ten polymorphisms utilized have already been associated with endurance athlete status and/or endurance-related phenotypes in discrete studies (Buemann et al. 2001; Halsall et al. 2001; Lucia et al. 2005; Tang et al. 2005; Ahmetov et al. 2006; Prior et al. 2006; Ahmetov et al. 2007a, b, 2008a, b, c; Eynon et al. 2009), so our data are

partially replicating these previous findings. We have here identified three other polymorphisms that include 'endurance' alleles that combine with the other seven to collectively associate with endurance phenotypes in humans. Collectively, we have shown that the *NFATC4* Gly160, *PPARA* rs4253778 G, *PPARD* rs2016520 C, *PPARGC1A* Gly482, *PPARGC1B* 203Pro, *PPP3R1* promoter 5I, *TFAM* 12Thr, *UCP2* 55Val, *UCP3* rs1800849 T and *VEGFA* rs2010963 C alleles are likely to be advantageous for performance in endurance sports, and having nine or more of these alleles increases the likelihood of becoming an elite endurance athlete. This conclusion is based on the differences in allelic frequencies between endurance-oriented athletes and controls, and in the proportions of subjects with high numbers of 'endurance' alleles between athletes of different competitive standard. These differences between athletes within the same event category but with different levels of achievement may be explained by selection pressure - i.e. those endurance-oriented athletes who possess a higher number of 'endurance' alleles are more likely to become elite. Of those listed, five alleles (*NFATC4* Gly160, *PPARA* rs4253778 G, *PPARGC1A* Gly482, *PPP3R1* promoter 5I and *UCP2* 55Val) are frequently observed in European populations (frequency: 37-91%), the others less frequently (5-25%). Three alleles, namely *NFATC4* Gly160, *PPARGC1B* 203Pro and *TFAM* 12Thr are novel genetic markers associated with endurance athlete status. *NFATC4* and *TFAM* code for transcription factors, whilst *PPARGC1B* codes for a transcription co-activator. These proteins interact with PPARα, PPARGC1α and PPP3R1, and are involved in glucose and lipid metabolism, mitochondrial biogenesis, formation of muscle fibers and regulation of cardiac hypertrophy (Allen et al. 2001; Poirier et al. 2003; Ekstrand et al. 2004; Yang et al. 2006; Arany et al. 2007; Ling et al. 2007; Alvarez et al. 2008).

Independent of the case-control study regarding allele frequencies of genes, we also investigated intermediate phenotypes (maximal oxygen consumption and muscle fiber composition) that are known to be related to endurance performance to begin to create a chain of evidence linking the polymorphisms we studied to success in endurance-oriented sports. Indeed, the genes *NFATC4*, *PPARA*, *PPARD*, *PPARGC1A*, *PPARGC1B*, *PPP3R1*, *TFAM*, *UCP2*, *UCP3* and *VEGFA* have been shown to influence muscle fiber composition and/or skeletal muscle and myocardium metabolism (see Table 1). Maximal rate of oxygen consumption is dependent on a variety of factors, such as maximal cardiac output and the ability of cardiac and skeletal muscle to take up and use oxygen from the local capillary bed. Accordingly, we found the number of 'endurance' alleles possessed by rowers of national competitive standard to be positively correlated with maximal oxygen consumption values - a widely-used laboratory parameter related to proficiency in endurance-related activities. Within skeletal muscle itself, fiber type proportion is a useful marker of skeletal muscle functional properties, and a high proportion of slow-twitch fibers is related to high mitochondrial volume, high oxidative capacity

and high fatigue resistance. Accordingly, we found that the number of 'endurance' alleles possessed positively correlated with the proportion of *m. vastus lateralis* slow-twitch fibers in men.

To date, few studies have sought to define or quantify the impact of genotype combinations that influence human physical performance and have used only 2-7 genetic markers at one time (Williams et al. 2004; Saunders et al. 2006; Gonzalez-Freire et al. 2008; Muniesa et al. 2008; Gómez-Gallego et al. 2009; Ruiz et al. 2009). For instance, Williams et al. (2004) have shown evidence for an interaction between the *BDKRB2* (bradykinin receptor B2) –9/+9 and *ACE* I/D polymorphisms in 115 British subjects, with individuals who were carriers of the *ACE* II + *BDRRB2* –9/–9 genotype combination having the highest efficiency of muscular contraction. Furthermore, the *ACE*(I)/*BDRRB2*(–9) ("high kinin receptor activity") haplotype was significantly associated with the distance of the preferred endurance event among elite British athletes ($P = 0.003$). Similarly, Saunders et al. (2006) found that the *NOS3* (nitric oxide synthase 3) Glu298 allele combined with a *BDKRB2* –9/–9 genotype was over-represented in the fastest-finishing Ironman triathletes (28.6%) compared with controls (17.3%; $P = 0.028$). Gómez-Gallego *et al.* (2009) have shown that professional road cyclists with the most strength/power oriented genotype combination, namely *ACE* DD + *ACTN3* (α-actininin-3) RR/RX, had higher respiratory compensation threshold values than those with the intermediate combinations (II + RX/RR, $P = 0.036$; and DD + XX, $P = 0.0004$) but similar to those with the II + XX genotype combination.

Recently, Ruiz *et al.* (2009) analysed seven genetic polymorphisms (*ACE*, *ACTN3*, *AMPD1*, *CKMM*, *HFE*, *GDF8* and *PPARGC1A*) in 46 world-class endurance athletes and 123 controls. Using the model developed by Williams and Folland (2008), they determined that the mean 'total genotype score' (TGS, from the accumulated combination of the seven polymorphisms, with a maximum value of '100' for the theoretically optimal polygenic score) was higher in athletes (70.2 ± 15.6) than in controls (62.4 ± 11.5) and also higher than predicted for the total Spanish population (60.8 ± 12.1), suggesting an overall more 'favorable' polygenic profile in the athlete group. The current study makes a further step forward with regards to a polygenic approach.

Our findings confirm the polygenic nature of a complex trait, such as elite endurance athlete status as proposed recently (Williams and Folland 2008; Ruiz et al. 2009). They also suggest that genetic science has advanced to the point where diverse loci of influence can be identified. The application of rapidly-advancing technologies at steadily-reducing cost (Metzker 2005; Mardis 2006; Eid et al. 2008) is likely to help the identification of many more such alleles and, in due course, allow gene-gene interaction to be studied. A strategy that directly considers gene-gene interactions will provide an opportunity for modeling interaction of differing biological systems, perhaps especially

when applied to prospective training studies. Such findings also have implications for the study of polygenic diseases, where multiple genetic and environmental influences may be at play.

Our findings may also have broader implications. Although certain polymorphisms have been associated with high levels of achievement in certain sporting disciplines, none (with the possible exception of the *ACTN3* null mutation and its relationship to sprint ability (MacArthur and North 2007)) has been of sufficient influence to be used in the selection of athletes. Whilst our data remain short of allowing this to happen with high confidence, they make clear that large-scale screening may have potential to identify 'rarer' allelic combinations which are strongly associated with specific training potential or 'latent ability'. Currently, 'performance tests' (such as time to run a given distance) or traditional laboratory tests (such as assessment of VO_{2max}, muscle strength, or anthropometry) are used to help identify young athletes with appropriate physiological potential, and to guide them into suitable training and competition. Such tests may, in the future, be augmented (or, in selected cases, superseded) by assessment of polygenic profile.

Our study does have limitations. Whilst the functional properties of *PPARD* rs2016520, *PPARGC1A* rs8192678, *PPARGC1B* rs7732671, *PPP3R1* promoter 5I/5D, *UCP2* rs660339, *UCP3* rs1800849 and *VEGFA* rs2010963 have been proposed (Schrauwen et al. 1999; Watson et al. 2000; Buemann et al. 2001; Skogsberg et al. 2003; Ling et al. 2004; Tang et al. 2005; Ling et al. 2007), the paucity of functional data relating to the *PPARA* rs4253778, *NFATC4* rs2229309 and *TFAM* rs1937 alleles needs to be addressed with further *in vitro* studies. Furthermore, extension to, and replication within groups of differing geographic ancestry is needed to translate these findings more broadly. Population stratification is a potential concern, as the controls were recruited from four specific regions, while the athletes originated from a more widely distributed area – i.e. across Russia. However, because there were no differences in genotype and allele frequencies between the citizens of the different regions (control group), it is unlikely that the associations observed are the result of population stratification in the current study. Nevertheless, recruiting the controls from four major population centers of Russia (as opposed to just one) ensured the control group was broadly comparable to and therefore representative of the wide population from which the athletes were also drawn.

In addition, our study is limited to 15 common polymorphisms which were primarily selected because of previously reported associations with various aspects of metabolism, and 10 of those were subsequently included in the main analyses. There are already other genetic variants that have been reported to show associations with aspects of endurance (Bray et al. 2009), and we strongly suspect that many additional common polymorphisms, and probably rare mutations as well, will be shown to be associated with endurance performance or an endurance phenotype in due course. Thus, we suspect

that the 10 polymorphisms we have used constitute only a fraction of the genetic factors that influence human endurance (Williams and Folland 2008). However, looking to the future, when a more complete analysis can be conducted that does in fact include the majority of polymorphisms that contribute to the variability in human physical endurance, the power of such an approach as a practical tool for sports coaches will be clear. Interestingly, there may be other areas of activity outside of the sporting world where highly competitive selection procedures routinely occur, such as military organizations, which may be receptive to such a tool that could increase the ability to predict physical performance potential.

In conclusion, our findings confirm the polygenic nature of endurance performance, a classic complex trait, and demonstrate that the likelihood of becoming an elite endurance athlete depends on the number of endurance-related alleles an individual possesses.

SUPPLEMENTARY MATERIAL

See attached file.

FUNDING

This work was supported by grants from the Federal Agency for Physical Culture and Sport of the Russian Federation and the Ministry of Education and Science of the Russian Federation (contract number 02.522.11.2004).

ACKNOWLEDGEMENTS

The authors thank A. Komkova, J. Shihova, A. Druzhevskaya and S. Khalchitsky (St Petersburg Research Institute of Physical Culture, St Petersburg) for their contributions to sample collection, genotyping and data management. We are also thankful to S. Missina for help with rowing spiroergometry, to B. Shenkman and P. Tarakin (SSC RF Institute for Biomedical Problems of the Russian Academy of Sciences, Moscow) for technical assistance in immunohistochemistry.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

REFERENCES

Abernethy PJ, Thayer R, Taylor AW (1990) Acute and chronic responses of skeletal muscle to endurance and sprint exercise. A review. Sports Med 10: 365-89.

Ahmetov II, Astratenkova IV, Rogozkin VA (2007a) Association of a *PPARD* polymorphism with human physical performance. Mol Biol 41:776-780.

Ahmetov II, Khakimullina AM, Popov DV, Missina SS, Vinogradova OL, Rogozkin VA (2008a) Polymorphism of the vascular endothelial growth factor gene (*VEGF*) and aerobic performance in athletes. Hum Physiol 34:477-481.

Ahmetov II, Linde EV, Shikhova YuV, Popov, DV, Missina SS, Vinogradova OL, Rogozkin VA (2008b) The influence of calcineurin gene polymorphism on morphofunctional characteristics of cardiovascular system of athletes. Ross Fiziol Zh Im I M Sechenova 94:915-922.

Ahmetov II, Mozhayskaya IA, Flavell DM, Astratenkova IV, Komkova AI, Lyubaeva EV, Tarakin PP, Shenkman BS, Vdovina AB, Netreba AI et al (2006) PPARα gene variation and physical performance in Russian athletes. Eur J Appl Physiol 97:103-108.

Ahmetov II, Popov DV, Astratenkova IV, Druzhevskaya AM, Missina SS, Vinogradova OL, Rogozkin VA (2008c) The use of molecular genetic methods for prognosis of aerobic and anaerobic performance in athletes. Hum Physiol 34:338-342.

Ahmetov II, Popov DV, Mozhayskaya IA, Missina SS, Astratenkova IV, Vinogradova OL, Rogozkin VA (2007b) Association of regulatory genes polymorphisms with aerobic and anaerobic performance of athletes. Ross Fiziol Zh Im I M Sechenova 93:837-43.

Ahmetov II, Rogozkin VA (2009) Genes, athlete status and training – An overview. In: Genetics and Sports, ed. Collins M. Med Sport Sci. Basel, Karger 54:43-71.

Allen DL, Sartorius CA, Sycuro LK, Leinwand LA (2001) Different Pathways Regulate Expression of the Skeletal Myosin Heavy Chain. Genes 276:43524-43533.

Alvarez V, Corao AI, Alonso-Montes C, Sánchez-Ferrero E, De Mena L, Morales B, García-Castro M, Coto E (2008) Mitochondrial Transcription Factor A (*TFAM*) Gene Variation and Risk of Late-Onset Alzheimer's Disease. J Alzheimers Dis 13:275-280.

Arany Z, Lebrasseur N, Morris C, Smith E, Yang W, Ma Y, Chin S, Spiegelman BM (2007) The Transcriptional Coactivator PGC-1β Drives the Formation of Oxidative Type IIX Fibers in Skeletal Muscle. Cell Metab 5:35-46.

Bassett DR Jr, Howley ET (2000) Limiting factors for maximum oxygen uptake and determinants of endurance performance. Med Sci Sports Exerc 32:70-84.

Bouchard C, An P, Rice T, Skinner JS, Wilmore JH, Gagnon J, Pérusse L, Leon AS, Rao DC (1999) Familial aggregation of VO_{2max} response to exercise training: results from the HERITAGE Family Study. J Appl Physiol 87:1003-1008.

Bouchard C, Daw EW, Rice T, Pérusse L, Gagnon J, Province MA, Leon AS, Rao DC, Skinner JS, Wilmore JH (1998) Familial resemblance for VO_{2max} in the sedentary state: the HERITAGE Family Study. Med Sci Sports Exerc 30:252-258.

Brand MD, Esteves TC (2005) Physiological functions of the mitochondrial uncoupling proteins UCP2 and UCP3. Cell Metab 2:85-93.

Bray MS, Hagberg JM, Perusse L, Rankinen T, Roth SM, Wolfarth B, Bouchard C (2009) The Human Gene Map for Performance and Health-Related Fitness Phenotypes: The 2006-2007 Update. Med Sci Sports Exerc 41:35-73.

Brutsaert TD, Parra EJ (2006) What makes a champion? Explaining variation in human athletic performance. Respir Physiol Neurobiol 151:109-123.

Buemann B, Schierning B, Toubro S, Bibby BM, Sørensen T, Dalgaard L, Pedersen O, Astrup A (2001) The association between the val/ala-55 polymorphism of the uncoupling protein 2 gene and exercise efficiency. Int J Obes Relat Metab Disord 25:467-471.

Chin ER, Olson EN, Richardson JA, Yang Q, Humphries C, Shelton JM, Wu H, Zhu W, Bassel-Duby R, Williams RS (1998) A calcineurin-dependent pathway controls skeletal muscle fiber type. Genes Dev 12:2499-2509.

De Moor MH, Spector TD, Cherkas LF, Falchi M, Hottenga JJ, Boomsma DI, De Geus EJ (2007) Genome-wide linkage scan for athlete status in 700 British female DZ twin pairs. Twin Res Hum Genet 10:812-820.

Dzau VJ (1988) Tissue renin-angiotensin system: physiologic and pharmacologic implications. Introduction. Circulation 77:I1–13.

Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B et al (2009) Real-time DNA sequencing from single polymerase molecules. Science 323:133-138.

Ekstrand MI, Falkenberg M, Rantanen A, Park CB, Gaspari M, Hultenby K, Rustin P, Gustafsson CM, Larsson NG (2004) Mitochondrial transcription factor A regulates mtDNA copy number in mammals. Hum Mol Genet 13:935-944.

Eynon N, Meckel Y, Sagiv M, Yamin C, Amir R, Sagiv M, Goldhammer E, Duarte JA, Oliveira J (2009) Do PPARGC1A and PPARalpha polymorphisms influence sprint or endurance phenotypes? Scand J Med Sci Sports DOI:10.1111/j.1600-0838.2009.00930.x.

Ferrara N (2001) Role of vascular endothelial growth factor in regulation of physiological angiogenesis. Am J Physiol Cell Physiol 280:1358-1366.

Flück M (2006) Functional, structural and molecular plasticity of mammalian skeletal muscle in response to exercise stimuli. J Exp Biol 209(Pt 12):2239-48.

Gayagay G, Yu B, Hambly B, Boston T, Hahn A, Celermajer DS, Trent RJ (1998) Elite endurance athletes and the *ACE* I allele – the role of genes in athletic performance. Hum Genet 103:48-50.

Gómez-Gallego F, Santiago C, González-Freire M, Muniesa CA, Fernández Del Valle M, Pérez M, Foster C, Lucia A (2009) Endurance performance: genes or gene combinations? Int J Sports Med 30:66-72.

Gonzalez-Freire M, Santiago C, Verde Z, Lao JI, Oiivan J, Gómez-Gallego F, Lucia A (2008) Unique among unique. Is it genetically determined? Br J Sports Med DOI:10.1136/bjsm.2008.049809.

Halsall DJ, Luan J, Saker P, Huxtable S, Farooqi IS, Keogh J, Wareham NJ, O'Rahilly S (2001) Uncoupling protein 3 genetic variants in human obesity: the c-55t promoter polymorphism is negatively correlated with body mass index in a UK Caucasian population. Int J Obes Relat Metab Disord 25:472-477.

Jamshidi Y, Montgomery HE, Hense H-W, Myerson SG, Torra IP, Staels B, World MJ, Doering A, Erdmann J, Hengstenberg C et al (2002) Peroxisome proliferators-activated receptor α gene regulates left ventricular growth in response to exercise and hypertension. Circulation 105:950-955.

Lefebvre P, Chinetti G, Fruchart JC, Staels B (2006) Sorting out the roles of PPARα in energy metabolism and vascular homeostasis. J Clin Invest 116:571-580.

Lin J, Wu H, Tarr PT, Zhang CY, Wu Z, Boss O, Michael LF, Puigserver P, Isotani E, Olson EN et al (2002) Transcriptional co-activator PGC-1α drives the formation of slow-twitch muscle fibres. Nature 418:797–801.

Ling C, Poulsen P, Carlsson E, Ridderstrale M, Almgren P, Wojtaszewski J, Beck-Nielsen H, Groop L, Vaag A (2004) Multiple environmental and genetic factors influence skeletal muscle PGC-1alpha and PGC-1beta gene expression in twins. J Clin Invest 114:1518-1526.

Ling C, Wegner L, Andersen G, Almgren P, Hansen T, Pedersen O, Groop L, Vaag A, Poulsen P (2007) Impact of the peroxisome proliferator activated receptor-γ coactivator-1β (PGC-1β) Ala203Pro polymorphism on in vivo metabolism, PGC-1β expression and fibre type composition in human skeletal muscle. Diabetologia 50:1615-1620.

Long YC, Glund S, Garcia-Roves PM, Zierath JR (2007) Calcineurin Regulates Skeletal Muscle Metabolism via Coordinated Changes in Gene Expression. J Biol Chem 282:1607-1614.

Lowenstein JM (1972) Ammonia production in muscle and other tissues: the purine nucleotide cycle. Physiol Rev 52:382–414.

Lucia A, Gómez-Gallego F, Barroso I, Rabadán M, Bandrés F, San Juan AF, Chicharro JL, Ekelund U, Brage S, Earnest CP et al (2005) *PPARGC1A* genotype (Gly482Ser) predicts exceptional endurance capacity in European men. J Appl Physiol 99:344-348.

MacArthur DG, North KN (2007) *ACTN3*: A genetic influence on muscle function and athletic performance. Exerc Sport Sci Rev 35:30-34.

Mardis ER (2006). Anticipating the 1,000 dollar genome. Genome Biol 7:112.

Metzker ML (2005). Emerging technologies in DNA sequencing. Genome Res 15:1767-1776.

Montgomery HE, Marshall R, Hemingway H, Myerson S, Clarkson P, Dollery C, Hayward M, Holliman DE, Jubb M, World M, et al (1998) Human gene for physical performance. Nature 393:221-222.

Moore RL (1998) Cellular adaptations of the heart muscle to exercise training. Ann Med 30(Suppl 1):46-53.

Muniesa CA, González-Freire M, Santiago C, Lao JI, Buxens A, Rubio JC, Martín MA, Arenas J, Gomez-Gallego F, Lucia A (2008) World-class performance in lightweight rowing: Is it genetically influenced? A comparison with cyclists, runners and non-athletes. Br J Sports Med DOI:10.1136/bjsm.2008.051680.

Narravula S, Colgan SP (2001) Hypoxia-inducible factor 1-mediated inhibition of peroxisome proliferator-activated receptor alpha expression during hypoxia. J Immunol 2001 166:7543-7548.

Poirier O, Nicaud V, McDonagh T, Dargie HJ, Desnos M, Dorent R, Roizès G, Schwartz K, Tiret L, Komajda M et al (2003) Polymorphisms of genes of the cardiac calcineurin pathway and cardiac hypertrophy. Eur J Hum Genet 11:659-664.

Prior SJ, Hagberg JM, Paton CM, Douglass LW, Brown MD, McLenithan JC, Roth SM (2006) DNA sequence variation in the promoter region of the VEGF gene impacts VEGF gene expression and maximal oxygen consumption. Am J Physiol Heart Circ Physiol 290:1848-1855.

Rico-Sanz J, Rankinen T, Joanisse DR, Leon AS, Skinner JS, Wilmore JH, Rao DC, Bouchard C; HERITAGE Family study (2003) Associations between cardiorespiratory responses to exercise and the C34T *AMPD1* gene polymorphism in the HERITAGE Family Study. Physiol Genomics 14:161-166.

Ruiz JR, Gómez-Gallego F, Santiago C, González-Freire M, Verde Z, Foster C, Lucia A (2009) Is there an optimum endurance polygenic profile? J Physiol 587(Pt 7):1527-1534.

Saunders CJ, Xenophontos SL, Cariolou MA, Anastassiades LC, Noakes TD, Collins M (2006) The bradykinin b2 receptor (*BDKRB2*) and endothelial nitric oxide synthase 3 (*NOS3*) genes and endurance performance during Ironman Triathlons. Hum Mol Genet 15:979-987.

Schrauwen P, Xia J, Walder K, Snitker S, Ravussin E (1999) A novel polymorphism in the proximal UCP3 promoter region: effect on skeletal muscle *UCP3* mRNA expression and obesity in male non-diabetic Pima Indians. Int J Obes Relat Metab Disord 23:1242-1245.

Semenza GL (2000) HIF-1: mediator of physiological and pathophysiological responses to hypoxia. J Appl Physiol 88:1474-1480.

Simoneau J-A, Bouchard C (1995) Genetic determinism of fiber type proportion in human skeletal muscle. FASEB J 9:1091-1095.

Semple RK, Chatterjee VK, O'Rahilly S (2006) PPAR gamma and human metabolic disease J Clin Invest 116:581-589.

Skogsberg J, Kannisto K, Cassel TN, Hamsten A, Eriksson P, Ehrenborg E (2003) Evidence That Peroxisome Proliferator-Activated Receptor Delta Influences Cholesterol Metabolism in Men. Arterioscler Thromb Vasc Biol 23:637-643.

St-Pierre J, Lin J, Krauss S, Tarr PT, Yang R, Newgard CB, Spiegelman BM (2003) Bioenergetic analysis of peroxisome proliferator-activated receptor gamma coactivators 1alpha and 1beta (PGC-1alpha and PGC-1beta) in muscle cells. J Biol Chem 278:26597-26603.

Tang W, Arnett DK, Devereux RB, Panagiotou D, Province MA, Miller MB, de Simone G, Gu C, Ferrell RE (2005) Identification of a novel 5-base pair deletion in calcineurin B (*PPP3R1*) promoter region and its association with left ventricular hypertrophy. Am Heart J 150:845-851.

Vanttinen M, Nuutila P, Kuulasmaa T, Pihlajamaki J, Hallsten K, Virtanen KA, Lautamaki R, Peltoniemi P, Takala T, Viljanen AP, Knuuti J, Laakso M (2005) Single nucleotide polymorphisms in the peroxisome proliferator-activated receptor delta gene are associated with skeletal muscle glucose uptake. Diabetes 54:3587-3591.

Wang D, Wang H, Guo Y, Ning W, Katkuri S, Wahli W, Desvergne B, Dey SK, DuBois RN (2006) Crosstalk between peroxisome proliferator-activated receptor delta and VEGF stimulates cancer progression. Proc Natl Acad Sci USA 103:19069-19074.

Wang YX, Zhang CL, Yu RT, Cho HK, Nelson MC, Bayuga-Ocampo CR, Ham J, Kang H, Evans RM (2004) Regulation of muscle fiber type and running endurance by PPARδ. PLoS Biol 2:294.

Watson CJ, Webb NJ, Bottomley MJ, Brenchley PE (2000) Identification of polymorphisms within the vascular endothelial growth factor (*VEGF*) gene: correlation with variation in VEGF protein production. Cytokine 12:1232-1235.

Williams AG, Dhamrait SS, Wootton PTE, Day SH, Hawe E, Payne JR, Myerson SG, World M, Budgett R, Humphries SE, Montgomery HE (2004) Bradykinin receptor gene variant and human physical performance. J Appl Physiol 96:938-942.

Williams AG, Folland JP (2008) Similarity of polygenic profiles limits the potential for elite human physical performance. J Physiol 586:113-121.

Yang TTC, Suk HY, Yang X, Olabisi O, Yu RY, Durand J, Jelicks LA, Kim JY, Scherer PE, Wang Y et al (2006) Role of Transcription Factor NFAT in Glucose and Insulin Homeostasis. Mol Cell Biol 26:7372–7387.

Table 1. Candidate genes for endurance performance; their full names, functions of gene products, associated phenotypes and interactions

Gene	Full name	Functions, associated phenotypes and interactions
ACE	Angiotensin I converting enzyme	Regulates circulatory homeostasis through the synthesis of vasoconstrictor angiotensin II and the degradation of vasodilator kinins (Dzau 1988).
AMPD1	Adenosine monophosphate deaminase 1	Regulates muscle energy metabolism by catalyzing the deamination of adenosine monophosphate to inosine monophosphate (Lowenstein 1972; Rico-Sanz et al. 2003).
HIF1A	Hypoxia inducible factor 1, α subunit	Regulates the transcription of numerous genes in response to hypoxic stimuli. Genes responsive to HIF1 are involved in the processes of erythropoiesis, angiogenesis, and metabolism and include those encoding erythropoietin, VEGF, PPARα and glycolytic enzymes (Semenza 2000; Narravula and Colgan 2001).
NFATC4	Nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 4	Regulates cardiac hypertrophy, glucose and lipid metabolism, expression of the skeletal myosin heavy chain genes; regulates expression of *PPARG* (Allen et al. 2001; Poirier et al. 2003; Yang et al. 2006).
PPARA	Peroxisome proliferator-activated receptor α	Regulates liver, heart and skeletal muscle lipid metabolism, glucose homeostasis, mitochondrial biogenesis, cardiac hypertrophy, expression of *UCP2* and *UCP3* genes (Lefebvre et al. 2006).
PPARD	Peroxisome proliferator-activated receptor δ	Regulates fatty acid β-oxidation, glucose utilization, mitochondrial biogenesis, angiogenesis, muscle fiber type, expression of *PPARGC1A*, *UCP2*, *UCP3* and *VEGFA* genes (Wang et al. 2004, 2006).
PPARG	Peroxisome proliferator-activated receptor γ	Plays a critical physiological role as a central transcriptional regulator of adipogenic and lipogenic programs, insulin sensitivity and glucose homeostasis (Semple et al. 2006).
PPARGC1A	Peroxisome proliferator-activated receptor γ coactivator 1 α	Regulates fatty acid oxidation, glucose utilization, mitochondrial biogenesis, thermogenesis, angiogenesis, formation of muscle fibers; co-activates PPARα, PPARδ; regulates *TFAM* and *VEGFA* expression (Lin et al. 2002; St-Pierre et al. 2003).
PPARGC1B	Peroxisome proliferator-activated receptor γ coactivator 1 β	Regulates fatty acid oxidation, mitochondrial biogenesis, formation of muscle fibers; co-activates PPARα and PPARγ (St-Pierre et al. 2003; Arany et al. 2007; Ling et al. 2007).
PPP3R1	Protein phosphatase 3, regulatory subunit B, alpha isoform (calcineurin subunit B)	Confers calcium sensitivity, dephosphorylates and activates NFATC4. Regulates skeletal muscle and heart metabolism/hypertrophy, expression of *HIF1A*, *PPARA*, *PPARD*, *PPARGC1A* (Chin et al. 1998; Tang et al. 2005; Long et al. 2007).
TFAM	Mitochondrial transcription factor A	Involved in mitochondrial transcription regulation, proliferation of mitochondria and mitochondrial biogenesis (Ekstrand et al. 2004; Alvarez et al. 2008).
UCP2	Uncoupling protein 2	Uncouples oxidative phosphorylation from ATP synthesis; regulates lipid metabolism and energy expenditure (Buemann et al. 2001; Brand and Esteves 2005).
UCP3	Uncoupling protein 3	Uncouples oxidative phosphorylation from ATP synthesis; regulates lipid metabolism and energy expenditure, transports fatty acid anions out of mitochondria (Halsall et al. 2001; Brand and Esteves 2005).
VEGFA	Vascular endothelial growth factor A	Growth factor active in angiogenesis, vasculogenesis and endothelial cell growth; expression of VEGFA is regulated by HIF1 and calcineurin/NFAT signaling (Ferrara 2001; Prior et al. 2006).

Table 2. Characteristics of athletes from different sporting groups

Characteristics	Group				
	Long endurance	Middle endurance	Short endurance	Mixed	Power
Males	*n* = 178	*n* = 206	*n* = 73	*n* = 210	*n* = 331
Age (years)	24.5 ± 0.5	24.8 ± 0.3	24.2 ± 0.5	24.4 ± 0.3	26.0 ± 0.7
Height (cm)	176.9±0.5	186.9±0.7	178.5±0.7	174.3±0.8	177.1±0.6
Body mass (kg)	65.9±0.6	79.1±1.0	71.9±1.0	68.8±1.0	75.8±1.0
Females	*n* = 110	*n* = 84	*n* = 43	*n* = 38	*n* = 150
Age (years)	23.8 ± 0.6	24.8 ± 0.7	23.9 ± 1.1	24.2 ± 1.9	24.6 ± 0.6
Height (cm)	165.0 ± 0.6	177.3 ± 0.7	169.1 ± 0.9	163.5 ± 2.6	163.0 ± 0.9
Body mass (kg)	54.6 ± 0.8	71.7 ± 1.0	58.1 ± 0.9	54.0 ± 2.7	56.1 ± 1.1

Table 3. Minor allele frequencies of candidate genes in predominantly endurance-oriented athletes of national competitive standard and controls

Gene	Polymorphism	MA	Athletes		Controls		*P* value
			n	MAF	*n*	MAF	
ACE	Alu I/D (intron 16)	I	287	0.498	981	0.491	0.737
AMPD1	rs17602729 (nonsense C34T)	T	128	0.094	229	0.121	0.282
HIF1A	rs11549465 (missense C/T)	T	265	0.076	696	0.079	0.836
NFATC4	rs2229309 (missense **G**/C)	C	287	0.441	1132	0.561	2.5×10^{-7}
PPARA	rs4253778 (intron 7 C/**G**)	C	287	0.124	1132	0.164	0.018
PPARD	rs2016520 (5'UTR T/**C**)	C	287	0.19	1132	0.143	0.0060
PPARG	rs1801282 (missense C/G)	G	287	0.171	1132	0.153	0.292
PPARGC1A	rs8192678 (missense A/**G**)	A	287	0.256	1132	0.345	6.0×10^{-5}
PPARGC1B	rs7732671 (missense **C**/G)	C	287	0.08	1132	0.049	0.0040
PPP3R1	**5I**/5D (promoter)	5D	287	0.054	1132	0.087	0.0090
TFAM	rs1937 (missense G/**C**)	C	287	0.176	1132	0.091	6.1×10^{-9}
UCP2	rs660339 (missense C/**T**)	T	287	0.436	1132	0.367	0.0025
UCP3	rs1800849 (promoter C/**T**)	T	287	0.338	1132	0.242	3.0×10^{-6}
VEGFA	rs2010963 (promoter G/**C**)	C	287	0.305	1132	0.245	0.0030
VEGFA	rs699947 (promoter C/A)	C	287	0.481	1132	0.479	0.961

MA, minor allele; MAF, minor allele frequency. MA for *NFATC4* is C (Ala) and G (Gly) in predominantly endurance-oriented athletes (long and middle endurance groups) of national competitive standard and controls, respectively. 'Endurance' alleles (alleles associated with endurance athlete status in the present case-control study) are bold and underlined. *P* values < 0.05 indicate statistically significant differences between endurance athletes and controls by χ^2 test.

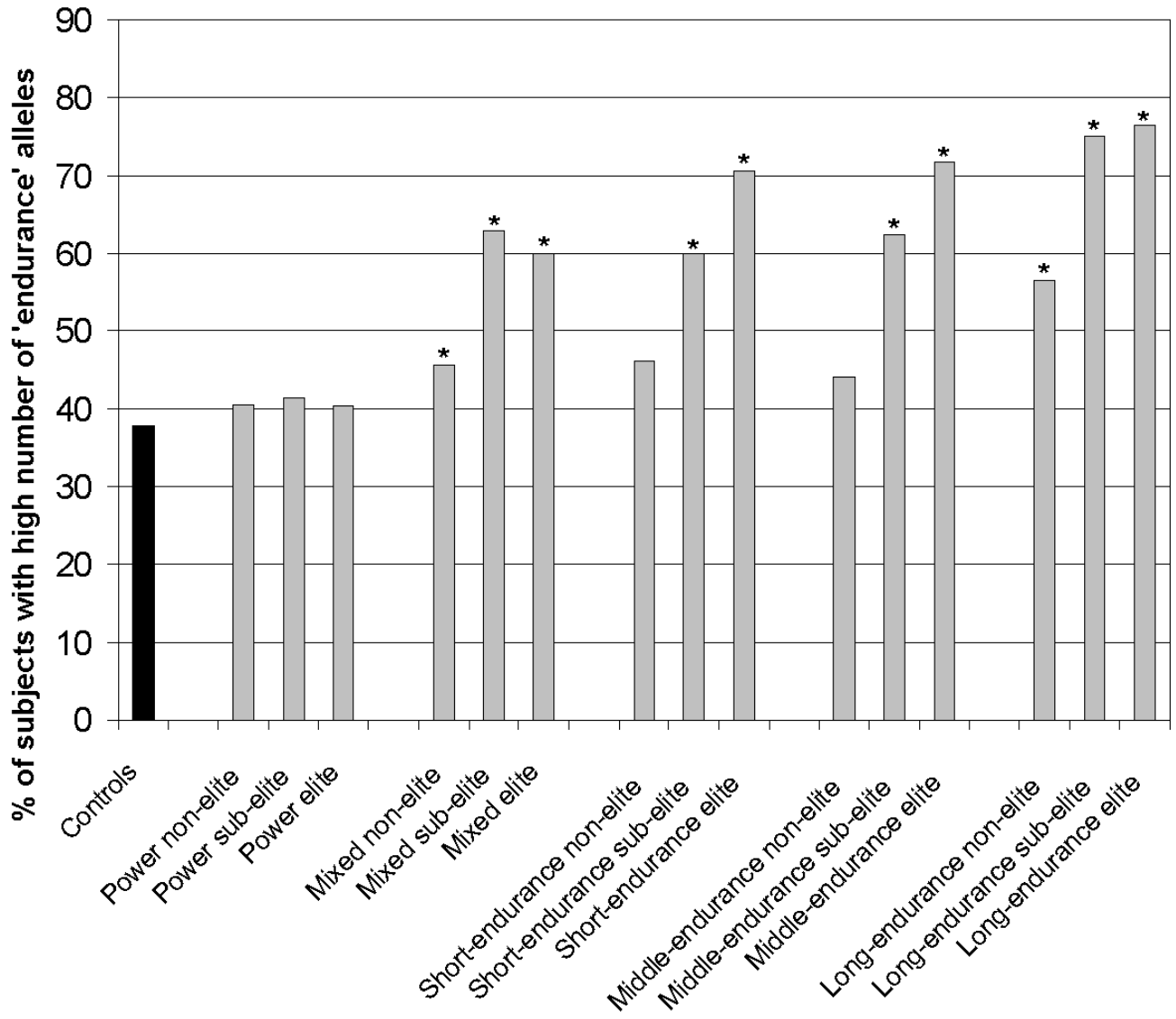

Figure 1. The combined impact of the 10 gene polymorphisms. The percentage of subjects with high (≥ 9) number of 'endurance' alleles is shown. Stratification of athletes was done by endurance/power orientation and competitive standard. *Significant difference ($P < 0.05$) from controls by χ^2 test. The proportion of subjects with a high number of 'endurance' alleles was significantly larger in the mixed group (non-elite: 45.6% (n = 193), P = 0.038; sub-elite: 62.9% (n = 35), P = 0.0026; elite: 60.0% (n = 20), P = 0.042), in the short-endurance group (non-elite: 46.2% (n = 39), P = 0.28; sub-elite: 60.0% (n = 60), $P = 5.6 \times 10^{-4}$; elite: 70.5% (n = 17), P = 0.0060), in the middle-endurance group (non-elite: 44.1% (n = 118), P = 0.18; sub-elite: 62.4% (n = 133), $P = 4.0 \times 10^{-8}$; elite: 71.7% (n = 39), $P = 1.8 \times 10^{-5}$) and in the long-endurance group (non-elite: 56.6% (n = 173), $P = 2.3 \times 10^{-6}$; sub-elite: 75.0% (n = 60), $P = 8.7 \times 10^{-9}$; elite: 76.4% (n = 55), $P = 1.0 \times 10^{-8}$) compared to controls (37.8% (n =1132)). On the contrary, the proportion of athletes with high number of 'endurance' alleles from the power group was not significantly different from controls (non-elite: 40.6% (n = 261); sub-elite: 41.4% (n = 116); elite: 40.4% (n = 104)).